Qwest's answers questions in Ken Lynch email

QUESTION:
"first: unique pids and description changes. pid uniqueness becomes an issue here simply because i often need to marry different files together. to ensure that i was merging correctly, i had been merging by state, product id, pid, dispatch status, and description. this worked well until the next round of data arrived on 6/28. the pid descriptions in the new data had changed in ways that were not substantive but made it impossible to merge the data. specifically, the attached spreadsheet shows that about 485 of the 700 or so descriptions changed (IA is used as an example here) from the 6/6 data to the 6/28 data. most of these can be accounted for by a missing comma before a trailing % sign, but i was a little uncomfortable making a global change to your data. other examples of changes are "Facil" became "Facility" and "Commit" became "Commitl". so to get around this, i just dropped the description from the merge and instead merged by state, prod id, pid and dispatch status. this works for most pids but there are some pids that are then not unique so i still got a lousy merge. for example, within any state, PO-2A-1, PO-2A-2, PO-2B-1 and PO-2B-2 are certainly different pids but have identical prod-ids, pid names, and dispatch status. anyway, either way, it was impossible to get a clean merge without picking through and changing the raw data."

ANSWER:
Use state, PID, Product, Dispatch and MSA-Zone as key fields for merging the files. The attached file (Comp_Jan02-May02.zip) provides monthly comparisons.

QUESTION:
"second: other questions we also had some other data related questions. for most of these, i suspect there is some documentation out there that explains them but i just haven't located it at this point. if you could point me in the right direction it would really save us some time.

"1. in the context of parity pids, are the z scores we see generated by permutation tests regardless of sample size or do you conduct a hypergeometric test for small samples? if the latter, what is a small sample (i.e., what is the cutoff)? if you could point me to some documentation on this issue i would really appreciate it."

ANSWER:
Qwest conducts permutation tests for interval-type variables if either the CLEC or retail sample size is 100 or less. Qwest performs an exact proportions test, based on a hypergeometric distribution, for binomial-type variables, if either the CLEC or retail sample size is 500 or less. (See attached document "Understanding Qwest's 272 Reports.doc".)

QUESTION:
"2. also in the context of parity pids, how do the trending reports and the aggregated spreadsheet data relate? for example, i assume that pid OP-3D and OP-3E from the trending reports are aggregated to form pid OP-3 (D/ND=blank) in the spreadsheets. using WA in 3/02 as an example, this seems to work. certainly the volumes and results add up. what about the z score and parity score? in the trending report, there is no qwest std deviation, and thus no z score, but there is a parity score. i guess i don't understand the parity score: i thought the parity score was essentially a function or scaling of the z score such that a parity score >0 indicated a statistical difference and a parity score <= 0 indicated no difference. this must not be the case because then how can there be a parity score when the z is N/A. further, the parity scores for the zone disaggregated pids are both less than 0 implying that there is no statistical difference between the results. yet when the data are aggregated, the z score is 1.8 which implies that there is a difference in the aggregate. i can understand why two marginal successes could be aggregated into a failure, but why is there a z score for the aggregated pid and no z scores for the disaggregated pids?"

ANSWER:
When a modified z-score is calculated, the parity score is calculated using the formula (modified z - critical z value) ÷ (critical z value). The critical z value is the one associated with the 95 percent confidence level, which is typically 1.645. However, when a permutation or exact proportions test is run, those tests determine the critical z value,[1] and the parity score is recalculated to take into account the results from those tests. This is done by finding p-values and converting the p-values to pseudo-z-scores using the standard normal distribution.

It is possible to generate a parity score even if a modified z-score can't be calculated. This is done with either a permutation test or exact proportions test. It is also possible that a z-score can be calculated for aggregated data even though it is not possible for the disaggregated data. This would happen if there is no variation in each of the disaggregated samples, but there is variation in the aggregated sample. Each of the disaggregated samples may have identical retail data in them (no variation), but the two samples may differ from one another. When they are aggregated, there is variation, and a standard deviation, standard error, and z-score can be calculated.

[1] Permutation and exact proportions test may determine a critical value that is different than 1.645, because the latter assumes a normal distribution, whereas the permutation and proportions tests make no such assumptions and address the actual distribution of the data to determine the critical value associated with the 95% confidence level.